NCBI

Transitioning from LocusLink to Entrez Gene
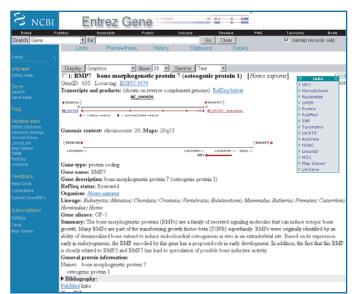
With the increasing sequencing and annotation of key genomes, having a gene-based view of the resultant information is useful. Entrez Gene has therefore been implemented to supply key connections in the nexus of map, sequence, expression, structure, function, citation, and homology data. Unique identifiers are assigned to any gene with defining sequence, genes with known map positions, and genes inferred from phenotypic information. These gene identifiers are tracked, and information is added when available. Access Entrez Gene from the Entrez Home Page or directly at:

www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

Entrez Gene can be considered as the successor to LocusLink, with the major differences being in greater scope (more of the genomes represented by NCBI Reference Sequences or RefSeqs) and in being integrated for indexing and query in NCBI's Entrez system. The Entrez Gene help document provides tips to ease LocusLink users into the current Entrez Gene resource. Sections involve interpreting the displays in a Gene record, as well as conversion of the LocusLink FTP file LL_tmpl to the Entrezgene.asn format for gene-specific information. The following section compares a sample LocusLink record for the human bone morphogenetic protein 7 (BMP7) to its counterpart entry in Entrez Gene.

The default format in Gene is the graphics display, which most closely resembles the traditional view of a LocusLink record. The array of colored boxes from LocusLink is replaced by the Links menu in Gene, which additionally supports many more links, such as to Books, GEO, UniSTS, Taxonomy, and others.

The Gene Transcripts and Products section is provided when a gene has been annotated on a genomic RefSeq; in other words when the intron/exon/coding region information is available in genomic coordinates. Each position of a gene product, when represented by a RefSeq RNA and/or protein, is provided relative to the genomic accession on which it is annotated. Each accession is an anchor to a menu allowing display of the sequence in several formats. Protein accessions also facilitate retrieval of specific BLink, CDD, or COG displays. This section is equivalent to the RNA-Genomic alignment available from the graphic at the top of a LocusLink entry. In the example of the Gene record for BMP7, NC_000020 is the accession number of the genomic contig that contains the gene. Clicking on the accession NC_000020 brings up a menu that allows navigation to a variety of display formats offered for that particular genomic range. The LocusLink record offers comparable genomic ranges from the various two-letter links adjacent to the NCBI contig records under the NCBI Reference Sequence(s) section.

The Entrez Gene record's General Gene Information section summarizes information contained in LocusLink's Function, Relationships and partially from Map Information sections. This section includes several subcategories of information, including Gene Ontology (GO), Homology, Phenotypes, Markers, Pathways and Relationships.

The remaining sections of an Entrez Gene record—NCBI Reference Sequences (RefSeqs), Related Sequences, and Additional Links—are equivalent to the corresponding entries in the LocusLink report. The first section describes the gene-specific NCBI reference sequences (RefSeqs) that have been established for the gene. In addition to enumerating the accessions and providing links to the appropriate Entrez sequence database, this section may also include descriptions of each transcript variant, accessions of the public sequences used to support any transcript, and a listing of computed domains in an encoded protein. The text provided in this section therefore supports retrieving gene records based on descriptions of conserved domains. The Related Sequences section lists nucleotide and protein accessions of sequences that are related to the gene, and provides links to the appropriate sequence record in Entrez nucleotide or protein. The Additional Links section provides a printable view of a subset of links to information both within and external to NCBI. Some of these links overlap those included in the Links menu. The intent of this section is to provide a printable report of, for example, MIM numbers, UniGene cluster numbers, and family-specific Web sites.

Where to find the content of LocusLink report page in Entrez Gene

LocusLink	Gene	Comments
Table of Contents	Not retained	
Alphabetic lists	Not retained	
Gene diagram	Transcripts and Products	Gene adds the function of Genomic context, to allow a quick view of nearby genes and links to their report pages.
Link to Evidence Viewer from Gene diagram	Evidence Viewer link in Links menu	The option to first see only the diagram of the alignment is not retained.
Button Links	Links menu	On the Gene Graphic/Default display, the number of links may be greater than in LocusLink.
Title bar with links to nomenclature source.	Initial text, with link to nomenclature source via LocusTag.	Links from LocusTag values may connect to an external database where official nomenclature has not been assigned.
Overview Section		
RefSeq Summary	Summary	not changed
Locus type	Combination of Gene type and evidence type (under development)	The text values are not equivalent. LocusLink's Locus type values are being subdivided into Gene type and Evidence type categories.
Protein names	General protein information	not changed
Alternate symbols	Gene aliases	not changed
Relationships Section		
Homology data	Links menu; HomoloGene	What is printed in LocusLink is still printed in Gene.
Related models	Related	not changed; limited to genomes being annotated by NCBI's pipeline.
Function Section		
GeneRIFS	GeneRIFs	Not in a function section; indented under Bibliography
GO annotation	General gene information: GeneOntology	organization changed, but not content
Phenotype	General gene information: Phenotypes	organization changed, but not content
Map Section		
Chromosome	Genomic context	not changed
Associated markers	General gene information: Sequence Tagged Site (Markers)	Entrez Gene added display of alternate marker names
Sequence Section		
RefSeq Subsection		
Category	RefSeq status	not changed
GenBank® source	Source sequence	not changed
Domain matches	Domains	CDD link also attached to the protein accession in **Transcripts and Products** and in Links menu.
BL (BLink)	BLink link attached to the protein accession in the Transcripts and Products section	The function was not changed, but the placement and visibility are different.
Variant name	After the protein accession in the Transcripts and Products section	content not changed
Annotation Subsection		
Genomic contig	Transcripts and Products	function retained
gb: Link to gene-specific subsequence	GENBANK view from source NC, NT, or NW accession	function retained as GENBANK option from the genomic accession-based menu.
sv: Link to graphic display of gene-specific subsequence	GRAPHICS view from source NC, NT, or NW accession	function retained as GRAPHIC option from the genomic accession-based menu.
mv: Map Viewer	Map Viewer in Links menu	function retained
ev: Evidence Viewer	Evidence Viewer in Links menu	function retained
mm: Model Maker	Model Maker in Links menu	function retained
strain or haplotype	not retained	
Related Sequences Subsection		
Accessions, type, and strain data	Related Sequences	content not changed
BL (BLink) from protein accessions	not retained	
Additional Link	Additional Links	